

* NOTICES *

JPO and NCIPJ are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

(57) [Claim(s)]

[Claim 1] The standard-pattern creation approach characterized by creating the standard pattern which makes the mixed continuous distribution with which the plurality of the vector output probability continuous distribution as which the vector output probability of said standard pattern is expressed in continuous distribution was mixed with weight in the creation approach of the standard pattern defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition a condition or the vector output probability of transition [claim 2] In the creation approach of the standard pattern defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition It learns using the speaker's voice data for every speaker about two or more speakers. The standard pattern which makes the mixed continuous distribution with which the plurality of the condition that the standard pattern by which the created vector output probability is expressed with continuous distribution corresponds, or the vector output probability continuous distribution of transition was mixed with weight a condition or the vector output probability of transition The standard-pattern creation approach characterized by creating [claim 3] In the creation approach of the standard pattern defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition It learns using the voice data uttered or recorded in a different environment. The standard pattern which makes the mixed continuous distribution with which the plurality of the condition that the standard pattern by which the created vector output probability is expressed with continuous distribution corresponds, or the vector output probability continuous distribution of transition was mixed with weight a condition or the vector output probability of transition The standard-pattern creation approach characterized by creating

[Translation done.]



*** NOTICES ***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]**[Industrial Application]**

This invention relates to the creation approach of the standard pattern used for pattern recognition, such as speech recognition.

[Description of the Prior Art]

In the field of pattern recognition, such as speech recognition, the approach using a probability model as a standard pattern for recognition attracts attention in recent years, and especially the hidden Markov model (it calls Following HMM) is widely used as a model which expresses a standard pattern in the field of speech recognition.

HMM is defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition, and recognizes by calculating each likelihood of HMM to an input configuration. The speech recognition by HMM is stated to publication "speech recognition by probability model" Nakagawa [Seiichi] work in detail.

The approach with which a parameter is repeated and updated, using the data for study as an approach of determining the parameter of the HMM model by which the vector output probability of each condition (or transition) is expressed with mixed continuous distribution is known from some initial value, such as a Baum-Welch algorithm. In this case, it is necessary to determine the initial value of parameters, such as the average of output probability distribution, for each [to mix] the distribution of every. It is the Institute of Electronics, Information and Communication Engineers voice study group data SP 89-48 which act on the parameter in distribution of the (b) single given by the (a) random numbers by obscuring with a random-number value as an approach of giving the initial value of these parameters ("examination of the Japanese phoneme recognition by the consecutive output distribution pattern HMM").

Which approach is learned.

As an approach of on the other hand determining a direct parameter from study data rather than asking by updating from a certain initial value (c) It asks for the cluster of the number of distribution clustered and mixed after carrying out the segmentation of the study data.

Parameters, such as the average, from the data of each class The approach of searching for is learned (). ["High Performance Connected Digit Recognition Using Hidden Markov Models",] [IEEE Transaction] on Acoustics, Speech, and Signal Processing, Vol.37, No.8, pp.1214-1224, and August 1989. Thus, it can also update with a Baum-Welch algorithm etc. by making the decided value into initial value.

[Problem(s) to be Solved by the Invention]

Although it is known that a setup of initial value is important in order to perform study efficiently when using the approach of updating a parameter repeatedly by study, possibility that will use a random number as shown in (a), or will take time amount by convergence of study in using the parameter in single distribution as shown in (b), and a convergence value will also turn into the whole local optimum value instead of an optimum value is high. On the other hand, although it was thought that the approach of (c) was converged by the count of a repeat small even when the study for renewal of a parameter is not necessarily needed and it uses as initial value of updating, there was a fault that needed to be calculated etc. for clustering and computational

This Page Blank (uspto)

complexity increased.

The purpose of this invention is to offer the standard-pattern creation approach which canceled such a fault.

[The means for solving a technical problem]

In the creation approach of a standard pattern that the 1st invention is defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition A vector output probability is characterized by creating the standard pattern which makes the mixed continuous distribution which mixed with weight the condition that two or more standard patterns expressed with continuous distribution correspond, or the vector output probability distribution of transition a condition or the vector output probability of transition.

In the creation approach of the standard pattern for speech recognition that the 2nd invention is defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition It learns using the speaker's voice data for every speaker about two or more speakers. The created vector output probability is characterized by creating the standard pattern which makes the mixed continuous distribution which mixed with weight the condition that the standard pattern expressed with continuous distribution corresponds, or the vector output probability distribution of transition a condition or the vector output probability of transition.

In the creation approach of the standard pattern for speech recognition that the 3rd invention is defined by the set of a condition, the transition probability between conditions, a condition, or the vector output probability of transition It learns for every environment using the voice data uttered or recorded in a different environment. The created vector output probability is characterized by creating the standard pattern which makes the mixed continuous distribution which mixed with weight the condition that the standard pattern expressed with continuous distribution corresponds, or the vector output probability distribution of transition a condition or the vector output probability of transition.

[Function]

According to this invention, the parameter of a standard pattern can be simply determined by compounding and searching for the vector output probability distribution expressed with mixed continuous distribution from the vector output probability distribution of two or more standard patterns [finishing / study / already]. Moreover, if the standard pattern used for composition is chosen appropriately, when using as initial parameters of study, such as the Baum-Welch method, it is expected that the probability which converges by the small count of study compared with the case where a random number determines an initial parameter etc., and is converged on a local optimum value also becomes small. Moreover, renewal of a parameter by study cannot be performed, but it can also use as it is.

If the standard pattern learned and created by composition using the speaker's voice data for every speaker about two or more speakers is used like the 2nd invention, a vector output probability can create simply the standard pattern for speaker independent speech recognition expressed with mixed consecutive output distribution.

If the standard pattern learned and created for every environment like the 3rd invention using the voice data uttered or recorded in an environment which is different in composition is used, a standard pattern strong against fluctuation of the environment where a vector output probability is expressed with mixed consecutive output distribution can be created simply.

[Example]

Fig. 1 is a block diagram for explaining the example which applied the 1st invention to the HMM model creation for speaker independent speech recognition. HMM model [from Speaker's B study data (2)] B (4) is created for HMM model A (3) from Speaker's A study data (1). As speakers A and B, it chooses one standard speaker at a time, for example from a male and a woman, and uses. Let a HMM model be the model of a form as shown in Fig. 2 . State transition probability a_{ii} and the output probability distribution $b_i(y)$ over a_{ii+1} ($a_{ii+1}=1$) and the output vector y are defined to each condition i . The state transition probability of Model A and output probability distribution are expressed as a_A , $b_A(y)$, etc., respectively. Supposing output vector

This Page Blank (uspto)

probability distribution is expressed with single Gaussian distribution $b_iA(y) = N(y, \mu_iA, \sigma_iA)$

$b_iB(y) = N(y, \mu_iB, \sigma_iB)$

It is expressed. Here, $N(y, \mu, \sigma)$ expresses the multi-dimension Gaussian distribution to which the mean vector is set to μ and it sets a covariance matrix to σ . From Model A and Model B, the HMM model C for speaker independent speech recognition (5) is created. State transition probability of Model C is set to $a_{ii}C$ and $a_{i+1}C$, and output probability distribution is set to b_iC . Output probability distribution presupposes that it is expressed with mixed Gaussian distribution with two following mixing.

$b_iC(y) = \lambda_1 N(y, \mu_1, \sigma_1)$

$+ \lambda_2 N(y, \mu_2, \sigma_2)$

At this time, each parameter of Model C is defined as follows.

$a_{ii}C = \{a_{ii}A + a_{ii}B\}/2$ $a_{i+1}C = \{a_{i+1}A + a_{i+1}B\}/2$ $\mu_1 = \mu_iA$, $\sigma_1 = \sigma_iA$ $\mu_2 = \mu_iB$,

$\sigma_2 = \sigma_iB$ $\lambda_1 = \lambda_2 = 1/2$ The model C created by doing in this way as the HMM model then for speaker independent speech recognition — it can also use — furthermore, many speakers' study data (6) — using — Baum-Welch — it can learn by law etc. and can also use as an early model for creating better model C' (7).

When what output probability distribution is expressed with mixed Gaussian distribution to as models A and B is prepared, Model C can be created similarly. In this case, the number of mixing of the output probability distribution of Model C becomes the sum of the number of mixing of the output probability distribution of Models A and B.

Next, one example of the 2nd invention is explained. By clustering the voice data of the a small number of vocabulary which many speakers uttered, a speaker is divided into M clusters and M speakers based on clusters are chosen from each cluster. About each M speakers, the HMM model by which output probability distribution is expressed according to single Gaussian distribution is learned and created based on the voice data of a complement to HMM study. The HMM model for speaker independent speech recognition is obtained from M created models by creating the HMM model with which the number of mixing makes the mixed Gaussian distribution of M output probability distribution like the example of the 1st invention. Since the data used for the clustering for choosing M speakers are good by a small number of data, computational complexity decreases compared with (c) of a Prior art.

One example of the 3rd invention is explained to the last. In the example of the 1st invention, if the model learned using the data uttered as how to choose Models A and B under the environment where a certain speakers differ (for example, a quiet environment and an environment with many noises) is used, a recognition model strong against environmental fluctuation as a model C can be created.

[Effect of the Invention]

As stated above, according to the 1st invention, using two or more standard patterns already learned, a vector output probability can determine easily the parameter of the standard pattern expressed with mixed continuous distribution, and can use for pattern recognition by study of the a small number of time which made initial value to remain as it is or this value. Moreover, according to the 2nd and 3rd invention, the object for unspecified speakers and a standard pattern strong against environmental fluctuation can be created simply, respectively.

[Translation done.]

This Page Blank (uspto)

* NOTICES *

JPO and NCIPJ are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

Fig. 1 is a block diagram for explaining the example which applied the 1st invention to the HMM model creation for speaker independent speech recognition,

Fig. 2 is drawing showing the form of the HMM model in an example.

- 1 Speaker's A study data
- 2 Speaker's B study data
- 3 HMM model A
- 4 HMM model B
- 5 HMM model C
- 6 An a large number speaker's study data
- 7 HMM model C'

[Translation done.]

This Page Blank (uspto)

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

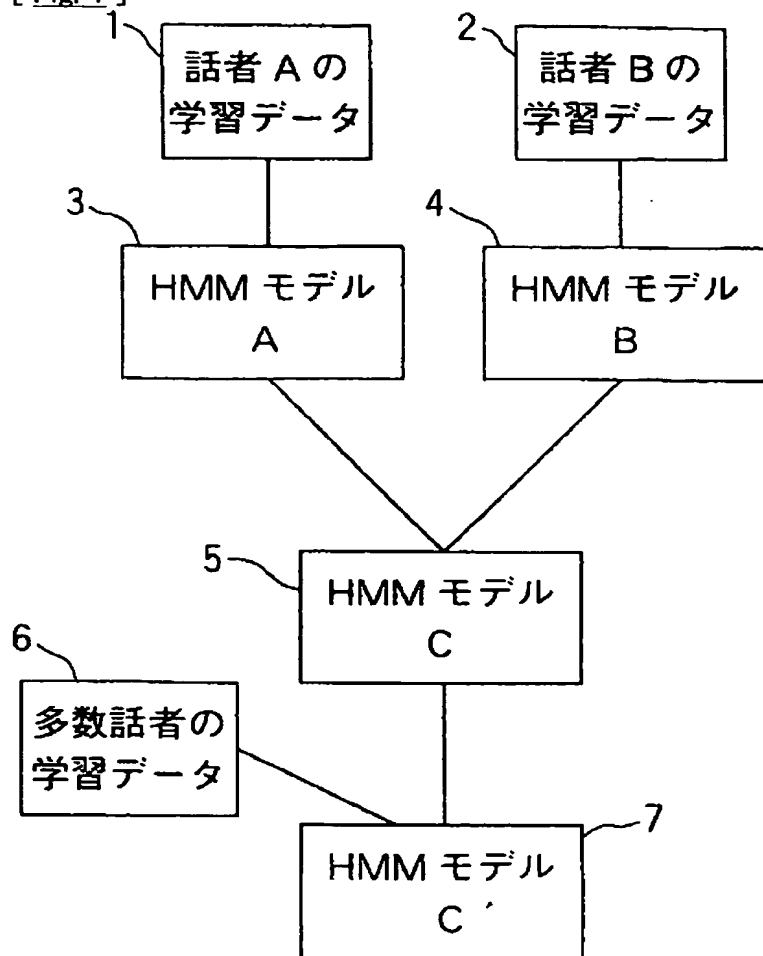
1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.*** shows the word which can not be translated.

3.In the drawings, any words are not translated.

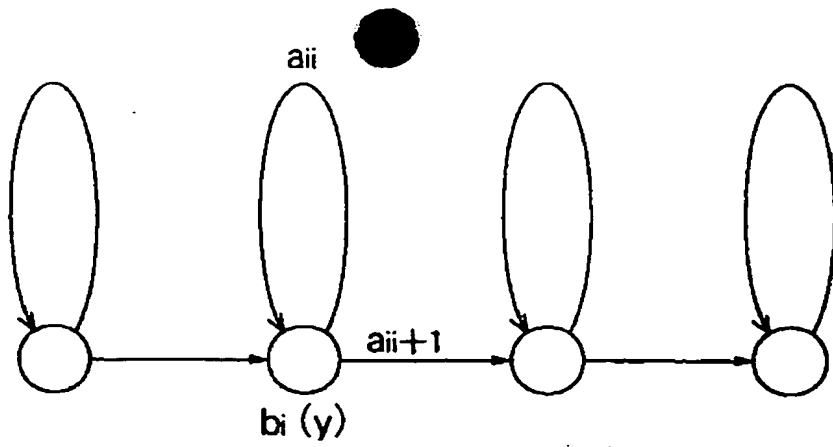
DRAWINGS

[Fig. 1]



[Fig. 2]

This Page Blank (uspto)



[Translation done.]

This Page Blank (uspto)

全項目

(19)【発行国】日本国特許庁(JP)
 (12)【公報種別】特許公報(B2)
 (11)【特許番号】特許第3251005号(P3251005)
 (24)【登録日】平成13年11月16日(2001. 11. 16)
 (45)【発行日】平成14年1月28日(2002. 1. 28)
 (54)【発明の名称】標準パターン作成方法
 (51)【国際特許分類第7版】

G10L 15/06

【FI】

G10L 3/00 521 C

【請求項の数】3

【全頁数】5

(21)【出願番号】特願平2-246863
 (22)【出願日】平成2年9月17日(1990. 9. 17)
 (65)【公開番号】特開平4-125599
 (43)【公開日】平成4年4月27日(1992. 4. 27)
 【審査請求日】平成6年11月30日(1994. 11. 30)
 【審判番号】平10-18000
 【審判請求日】平成10年11月12日(1998. 11. 12)
 (73)【特許権者】
 【識別番号】999999999
 【氏名又は名称】日本電気株式会社
 【住所又は居所】東京都港区芝5丁目7番1号
 (72)【発明者】
 【氏名】磯谷 亮輔
 【住所又は居所】東京都港区芝5丁目7番1号 日本電気株式会社内
 (74)【代理人】
 【識別番号】999999999
 【弁理士】
 【氏名又は名称】京本 直樹(外2名)
 【合議体】
 【審判長】谷川 洋
 【審判官】石川 伸一
 【審判官】小林 秀美
 (56)【参考文献】
 【文献】特公 昭63-67197(JP, B2)
 【文献】日本音響学会講演論文集 平成2年3月 2-3-1 P51~52
 【文献】中川著「確率モデルによる音声認識」(社団法人電子情報通信学会)昭和63年、第36~37行

(57)【特許請求の範囲】

【請求項1】状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とによって定義される標準パターンの作成方法において、前記標準パターンのベクトル出力確率が連続分布で表されるベクトル出力確率連続分布の複数が重み付きで混合された混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする標準パターン作成方法【請求項2】状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とに

よって定義される標準パターンの作成方法において、複数の話者について話者ごとにその話者の音声データを用いて学習して作成されたベクトル出力確率が連続分布で表される標準パターンの対応する状態あるいは遷移のベクトル出力確率連続分布の複数が重み付きで混合された混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする標準パターン作成方法【請求項3】状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とによって定義される標準パターンの作成方法において、異なる環境で発声あるいは収録した音声データを用いて学習して作成されたベクトル出力確率が連続分布で表される標準パターンの対応する状態あるいは遷移のベクトル出力確率連続分布の複数が重み付きで混合された混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする標準パターン作成方法

【発明の詳細な説明】

〔産業上の利用分野〕

本発明は、音声認識等パターン認識に用いられる標準パターンの作成方法に関する。

〔従来の技術〕

音声認識などパターン認識の分野で、認識用の標準パターンとして確率モデルを用いる方法が近年注目されており、特に隠れマルコフモデル(以下HMMと呼ぶ)は音声認識の分野で標準パターンを表すモデルとして広く用いられている。

HMMは状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率によって定義され、入力パターンに対する各HMMの尤度を計算することにより認識を行う。HMMによる音声認識については、刊行物「確率モデルによる音声認識」中川聖一著に詳しく述べられている。

各状態(あるいは遷移)のベクトル出力確率が混合連続分布で表されるHMMモデルのパラメータを決定する方法として、Baum-Welchアルゴリズムなど、ある初期値から学習用データを用いてパラメータを繰り返し更新する学習法が知られている。この場合、出力確率分布の平均値などのパラメータの初期値は、混合する各分布毎に決定する必要がある。これらのパラメータの初期値を与える方法としては、(a)乱数で与える(b)単一の分布の場合のパラメータに乱数値でばかし作用を行う(「連続出力分布型HMMによる日本語音韻認識の検討」電子情報通信学会音声研究会資料SP89-48)

などの方法が知られている。

一方、ある初期値から更新によって求めるのではなく学習データから直接パラメータを決定する方法として、(c)学習データをセグメンテーションしたあとクラスタリングを行って混合する分布数のクラスタを求め、各クラスのデータから平均値等のパラメータを求める方法が知られている(“High Performance Connected Digit Recognition Using Hidden Markov Models”, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol.37, No.8, pp.1214-1224, August 1989)。このようにして決められた値を初期値として、Baum-Welchアルゴリズムなどにより更新を行うこともできる。

〔発明が解決しようとする課題〕

学習により繰り返しパラメータを更新する方法を用いる場合、効率よく学習が行われるためには初期値の設定が重要であることが知られているが、(a)のように乱数を用いたり(b)のように単一の分布の場合のパラメータを用いるのでは、学習の収束までに時間がかかり、また収束値も全体の最適値ではなく局所的な最適値になる可能性が高い。一方(c)の方法は、パラメータ更新のための学習を必ずしも必要とせず、また、更新の初期値として用いる場合でも少ない繰り返し回数で収束すると考えられるが、クラスタリングのための計算などが必要で、計算量が多くなるという欠点があった。

本発明の目的は、このような欠点を解消した標準パターン作成方法を提供することにある。

〔課題を解決するための手段〕

第1の発明は、状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とによって定義される標準パターンの作成方法において、ベクトル出力確率が連続分布で表される複数の標準パターンの対応する状態あるいは遷移のベクトル出力確率分布を重み付きで混合した混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする。

第2の発明は、状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とによって定義される音声認識用の標準パターンの作成方法において、複数の話者について話者ごとにその話者の音声データを用いて学習して作成されたベクトル出力確率が連続分布で表される標

準パターンに対応する状態あるいは遷移のベクトル出力確率分布を重み付きで混合した混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする。第3の発明は、状態の集合と状態間の遷移確率と状態あるいは遷移のベクトル出力確率とによって定義される音声認識用の標準パターンの作成方法において、異なる環境で発声あるいは収録した音声データを用いて環境ごとに学習して作成されたベクトル出力確率が連続分布で表される標準パターンに対応する状態あるいは遷移のベクトル出力確率分布を重み付きで混合した混合連続分布を状態あるいは遷移のベクトル出力確率とする標準パターンを作成することを特徴とする。

〔作用〕

本発明によれば、混合連続分布で表されるベクトル出力確率分布を、すでに学習済みの複数の標準パターンのベクトル出力確率分布から合成して求めることにより、標準パターンのパラメータを簡易に決定することができる。また、合成に用いる標準パターンを適切に選べば、Baum-Welch法などの学習の初期パラメータとして用いる場合、乱数で初期パラメータを決定する場合などに比べ少ない学習回数で収束し、局所的な最適値に収束する確率も小さくなると期待される。また、学習によるパラメータ更新を行わずそのまま用いることもできる。

第2の発明のように、合成に複数の話者について話者ごとにその話者の音声データを用いて学習して作成された標準パターンを用いれば、ベクトル出力確率が混合連続出力分布で表される不特定話者音声認識用の標準パターンを簡易に作成することができる。

第3の発明のように、合成に異なる環境で発声あるいは収録した音声データを用いて環境ごとに学習して作成された標準パターンを用いれば、ベクトル出力確率が混合連続出力分布で表される環境の変動に強い標準パターンを簡易に作成することができる。

〔実施例〕

第1図は、第1の発明を不特定話者音声認識用のHMMモデル作成に適用した実施例を説明するためのブロック図である。話者Aの学習データ(1)からHMMモデルA(3)を、話者Bの学習データ(2)からHMMモデルB(4)を作成する。話者A,Bとしては、たとえば男性、女性から標準的な話者を1名ずつ選んで用いる。HMMモデルは第2図に示すような形のモデルとする。各状態*i*に対し、状態遷移確率 a_{ii}, a_{i+1} ($a_{ii} + a_{i+1} = 1$)と出力ベクトル y_i に対する出力確率分布 $b_i(y)$ が定められている。モデルAの状態遷移確率、出力確率分布を、それぞれ $a^A, b^A(y)$ などと表す。出力ベクトル確率分布が単一ガウス分布で表されたとすると、 $b_i^A(y) = N(y, \mu_i^A, \Sigma_i^A)$

$$b_i^B(y) = N(y, \mu_i^B, \Sigma_i^B)$$

と表される。ここで、 $N(y, \mu_i, \Sigma_i)$ は平均ベクトルを μ_i 、共分散行列を Σ_i とする多次元ガウス分布を表す。モデルAとモデルBから、不特定話者音声認識用のHMMモデルC(5)を作成する。モデルCの状態遷移確率を a_{ii}^C, a_{i+1}^C 、出力確率分布を b_i^C とする。出力確率分布が、次のような混合数2の混合ガウス分布で表されたとする。

$$b_i^C(y) = \lambda^1 N(y, \mu_i^1, \Sigma_i^1) + \lambda^2 N(y, \mu_i^2, \Sigma_i^2)$$

このとき、モデルCの各パラメータを次のように定める。

$$a_{ii}^C = [a_{ii}^A + a_{ii}^B] / 2 \quad a_{i+1}^C = [a_{i+1}^A + a_{i+1}^B] / 2 \quad \mu_i^1 = \mu_i^A, \Sigma_i^1 = \Sigma_i^A \quad \mu_i^2 = \mu_i^B, \Sigma_i^2 = \Sigma_i^B \quad \lambda^1 = \lambda^2 = 1/2$$

このようにして作成されたモデルCは、そのまま不特定話者音声認識用のHMMモデルとして用いることもでき、また、さらに多数の話者の学習データ(6)を用いてBaum-Welch法などで学習を行い、よりよいモデルC'(7)を作成するための初期モデルとして用いることもできる。

モデルA,Bとして出力確率分布が混合ガウス分布で表されるものが用意されている場合にも、同様にモデルCを作成することができる。この場合、モデルCの出力確率分布の混合数は、モデルA,Bの出力確率分布の混合数の和になる。

次に、第2の発明の一実施例について説明する。多数の話者が発声した少数語彙の音声データをクラスタリングすることにより話者をM個のクラスタに分け、各クラスタからクラスタ中心の話者M名を選ぶ。M名の各話者について、HMM学習に必要な量の音声データをもとに、出力確率分布が単一ガウス分布で表されるHMMモデルを学習して作成する。作成されたM個のモデルから、第1の発明の実施例と同様に混合数がMの混合ガウス分布を出力確率分布とするHMMモデルを作成することにより不特定話者音声認識用のHMMモデルが得られる。M名の話者を選ぶためのクラ

スタリングに用いるデータは少数のデータでよいので、従来の技術の(c)に比べ計算量は少なくなる。

最後に、第3の発明の一実施例について説明する。第1の発明の実施例において、モデルA,Bの選び方として、ある話者の異なる環境下(たとえば、静かな環境と雑音の多い環境)で発声したデータを用いて学習したモデルを用いれば、モデルCとして環境の変動に強い認識モデルを作成することができる。

〔発明の効果〕

以上述べたように、第1の発明によれば、すでに学習されている複数の標準パターンを用いて、ベクトル出力確率が混合連続分布で表される標準パターンのパラメータを簡単に決定することができる。そのまま、あるいはこの値を初期値とした少数回の学習でパターン認識に用いることができる。また、第2,第3の発明によれば、不特定話者用、環境の変動に強い標準パターンをそれぞれ簡易に作成することができる。

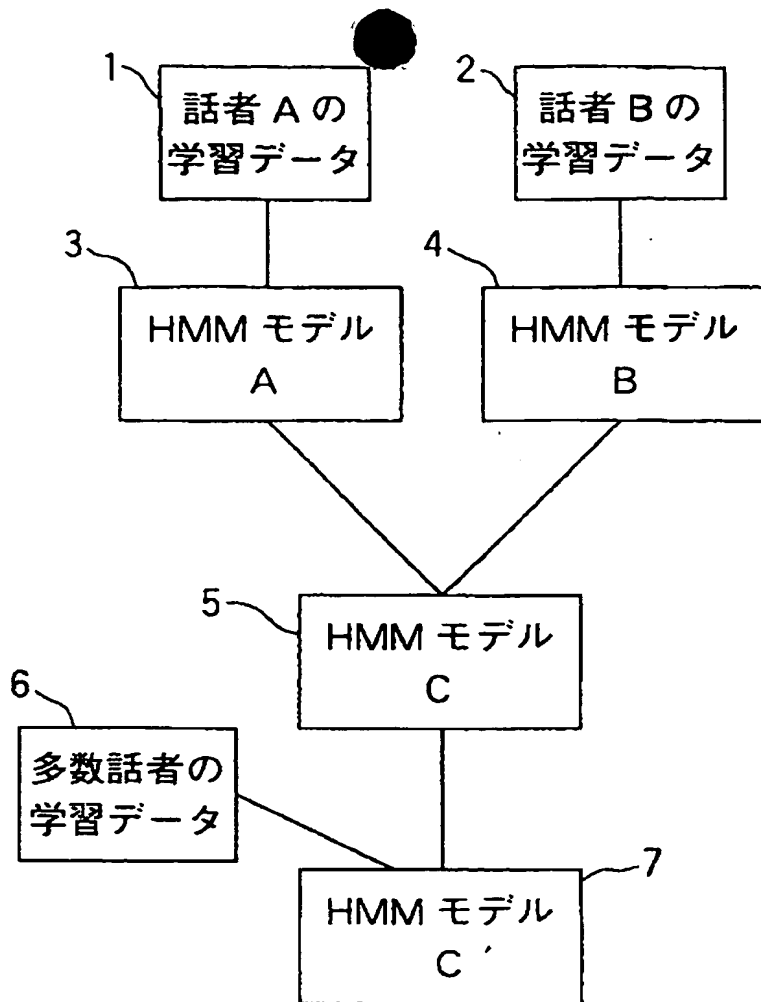
【図面の簡単な説明】

第1図は、第1の発明を不特定話者音声認識用のHMMモデル作成に適用した実施例を説明するためのブロック図、

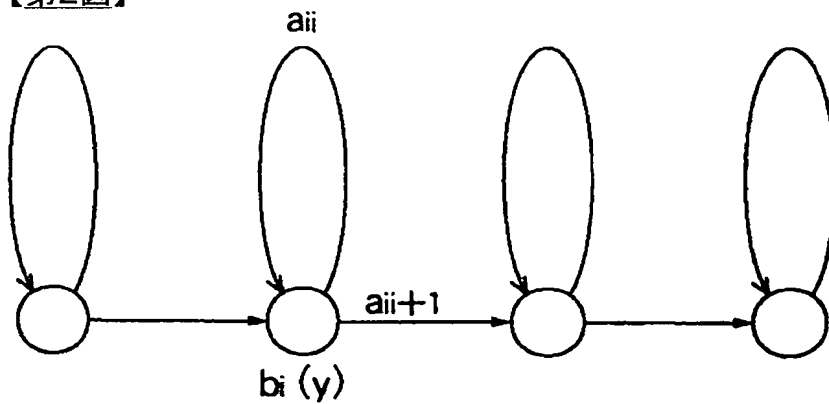
第2図は、実施例におけるHMMモデルの形を示す図である。

- 1……話者Aの学習データ
- 2……話者Bの学習データ
- 3……HMMモデルA
- 4……HMMモデルB
- 5……HMMモデルC
- 6……多数話者の学習データ
- 7……HMMモデルC'

【第1図】



【第2図】



This Page Blank (uspto)